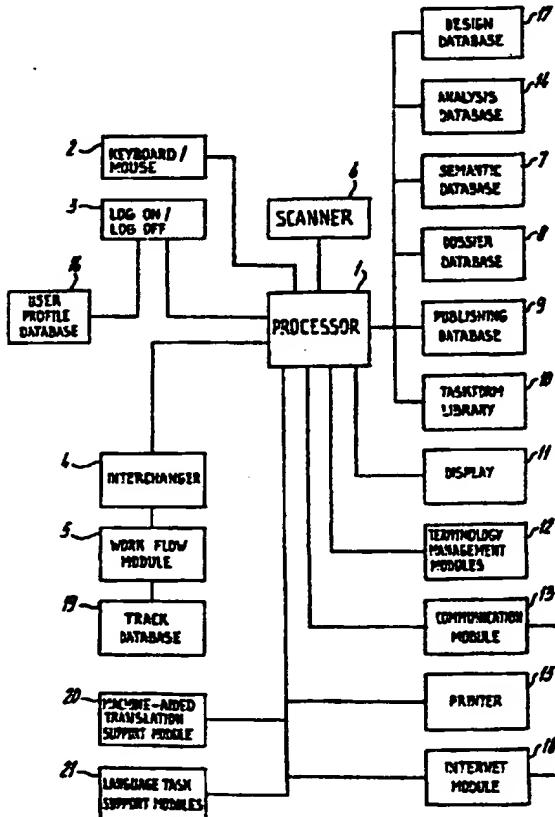




## INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification <sup>6</sup> :  G06F 17/28, 17/60		A1	(11) International Publication Number: <b>WO 98/01807</b>
			(43) International Publication Date: 15 January 1998 (15.01.98)
<p>(21) International Application Number: PCT/NL96/00273</p> <p>(22) International Filing Date: 3 July 1996 (03.07.96)</p> <p>(71) Applicant (<i>for all designated States except US</i>): POLYDOC N.V. [NL/NL]; 595-G A.J. Ernststraat, NL-1082 LD Amsterdam (NL).</p> <p>(72) Inventors; and</p> <p>(75) Inventors/Applicants (<i>for US only</i>): RUTTEN, Hubert, Joseph, Marie [NL/NL]; Spaulenstraat 4, NL-6245 BW Eijden (NL). VAN RUYSEVELT, Frank [NL/NL]; Pranquinestraat 32, NL-6217 KV Maastricht (NL). WILKINSON, David [GB/NL]; Rondenburglaan 113, NL-2135 KP Hoofddorp (NL).</p> <p>(74) Agents: BARENDREGT, F. et al.; Van Exter, Polak &amp; Charlouis, P.O. Box 3241, NL-2280 GE Rijswijk (NL).</p>		<p>(81) Designated States: AL, AM, AT, AU, AZ, BB, BG, BR, BY, CA, CH, CN, CZ, DE, DK, EE, ES, FI, GB, GE, HU, IL, IS, JP, KE, KG, KP, KR, KZ, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, TJ, TM, TR, TT, UA, UG, US, UZ, VN, ARIPO patent (KE, LS, MW, SD, SZ, UG), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, ML, MR, NE, SN, TD, TG).</p> <p>Published With international search report.</p>	
<p><b>(54) Title:</b> DOCUMENT PRODUCING SUPPORT SYSTEM</p> <p><b>(57) Abstract</b></p> <p>Document producing support system including: input means like a keyboard or a mouse; display means; a user profile database storing user profile; a logon/logout module connected to said user profile database for identifying roles of users based on said user profiles; a semantic database storing semantic elements; a dossier database for storing drafts of documents; a publishing database for storing final versions of documents; a work flow module arranged for at least controlling which roles of users are entitled to carry out specific activities and which activities may be carried out at a specific moment, as well as tracking which activities have already been carried out and how much process time has already been spent within any of the activities, activities being defined as sets of tasks allowed to be carried out by any individual user depending on his role; language technology modules for terminology control, machine-aided translation and language task support; a processor coupled to and arranged for controlling the elements mentioned above.</p>			



**FOR THE PURPOSES OF INFORMATION ONLY**

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslavia	TM	Turkmenistan
BF	Burkina Faso	GR	Greece	ML	Republic of Macedonia	TR	Turkey
BG	Bulgaria	HU	Hungary	MN	Mali	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MR	Mongolia	UA	Ukraine
BR	Brazil	IL	Israel	MW	Mauritania	UG	Uganda
BY	Belarus	IS	Iceland	MX	Malawi	US	United States of America
CA	Canada	IT	Italy	NB	Mexico	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NE	Niger	VN	Viet Nam
CG	Congo	KE	Kenya	NL	Netherlands	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NO	Norway	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's Republic of Korea	NZ	New Zealand		
CM	Cameroon	KR	Republic of Korea	PL	Poland		
CN	China	KZ	Kazakhstan	PT	Portugal		
CU	Cuba	LC	Saint Lucia	RO	Romania		
CZ	Czech Republic	LI	Liechtenstein	RU	Russian Federation		
DE	Germany	LK	Sri Lanka	SD	Sudan		
DK	Denmark	LR	Liberia	SE	Sweden		
EE	Estonia			SG	Singapore		

**Document producing support system**

The object of the present invention is to provide a system which is able to support users in the process of creating, maintaining, and publishing documents which have a clear, predictable structure by consistent use  
5 of language and flexible formats.

Therefore, the invention provides a document producing support system comprising:

- input means for receiving manual input from a user;
- 10 - display means;
- a user profile database storing user profiles;
- a logon/logoff module connected to said user profile database for identifying roles of users based on said user profiles;
- a semantic database storing semantic elements, i.e., pieces of data  
15 with pre-defined contents and meanings;
- a dossier database for storing drafts of documents;
- a publishing database for storing final versions of documents;
- a work flow module arranged for at least controlling which roles of  
20 users are entitled to carry out specific activities and which activities may be carried out at a specific moment, as well as tracking which activities have already been carried out and how much process time has already been spent within any of the activities, activities being defined as sets of tasks allowed to be carried out by any individual user depending on his role;
- 25 - language technology modules for terminology control, machine-aided translation and language task support;
- a processor coupled to and arranged for controlling said input means, said logon/logoff module, said work flow module, said display means, said language technology modules, and said semantic database, said dossier database and said publishing database.

Such a system provides an integrated software architecture for supporting language-intensive business processes involving extensive document production, wherein the linguistic structure of the text, their re-usability and their terminological consistency are dynamically enhanced during  
35 usage of the system by end-users in such a way that improvements of economy (speed of processing) and quality are attained.

The integrated software architecture for supporting language-intensive business processes defined by claim 1 structures these language-inten-

sive processes in such a way that end-users are enabled, at any given point of time, to work only on those specific activities which are relevant to them and which are logically possible according to a generic description of the process-type. Here, activities are defined as sets of tasks associated 5 with predetermined roles of users. An end-user may, e.g., be a manager, a technical writer, a translator, a claim-handler, a physician, etc. The end-user can only carry out those activities belonging to his or her role.

The integrated software architecture for supporting language-intensive business processes further breaks down the language-intensive activities 10 into language tasks and supports each of these with a variety of software applications, whereby the end-user is enabled to produce texts in such a way that the content is highly structured, the resulting text is highly re-usable, the terminology and extra-textural references consistent and that the style is adapted to the potential target group of the text.

15 The integrated software architecture for supporting language-intensive business processes is customizable in such a way that the process structure implemented for a particular client reflects both the generic process-type and the specific process organization preferred by the client.

When using the integrated software architecture for supporting language-intensive business processes the end user is enabled by the software 20 to build up a stock of computer files containing process-specific terminology, texts and text-components, extra-textual references and paralinguistic objects (i.e. drawings, images, audio-visuals) which can be used to generate an increasing variety of document-types and document-instances. Special 25 embodiments of the present invention are defined in the subclaims.

Below, the invention will be illustrated in detail with reference to some drawings, which are intended to illustrate the present invention and not to limit its scope.

Figure 1a shows a block diagram of the system according to the invention; 30

figures 1b and 1c show the content of some of the blocks of the block diagram of figure 1a;

figure 2 shows how a process of creating, maintaining, and publishing documents may be split up according to the invention, and

35 figure 3 shows a method which relates documents to business processes in order to evaluate their importance to the client's core business;

figure 4 shows a flow diagram of a document producing process.

As shown in figure 1a, the system according to the invention com-

prises a processor 1. The processor 1 is connected to input means, like a keyboard and/or mouse 2. The processor 1 is also connected to a logon/log-off module 3 which, in turn, is connected to a user profile database 16.

5 A workflow module 5 is connected to the processor 1 through an inter-changer 4, and to a tracking database 19.

Moreover, the processor 1 is connected to a scanner 6, a design database 17, an analysis database 14, a semantic database 7, a dossier database 8, a publishing database 9, a task form database 10, a display 11, terminology management modules 12, a communication module 13, a printer 15, an 10 internet connection module 18, a machine-aided translation support module 20, and language task support modules 21.

Figure 1b shows possible modules of the terminology management modules 12: a spelling checker module 12a, a terminology checker module 12b, a terminology database 12c, a monolingual custom dictionary module 12d, a 15 multilingual custom dictionary module 12e, a concordance module 12f, and a dynamic terminology collection module 12g.

Figure 1c shows possible modules belonging to the language task support modules 21: a reference insertion module, a module containing descriptions of procedures, a module containing materials descriptions, a 20 quotation insertion module and a summarizing module.

Below, terminology management modules 12, machine-aided translation support module 20, and language task support modules 21 will be together sometimes referred to as "language technology modules".

As shown in figure 2, in accordance with the invention a process is 25 subdivided into several sub-processes. Any of the sub-processes is subdivided into activities, each activity being relating to roles. Each activity comprises a set of tasks.

For instance, when the process relates to creating, maintaining, and publishing standard documents the sub-processes may for instance be:

- 30 a. intake
- b. collect information
- c. edit
- d. evaluate
- e. translate
- 35 f. authorize
- g. publish

As shown in figure 1a any sub-process is subdivided into activities related to roles of users. Any of the activities is supported by a so-

called "workbench" presented to the user by the processor 1 through the display 11. A workbench is the totality of functionality required to carry out the activity. A user accesses the functionality of the workbench, for instance, by means of a standard windows-type interface based on a menu bar, a tool bar, and a set of tabs.

The system controls who can carry out which activity in the process by means of the logon/logoff module 3 which identifies the user as having the necessary role (authorization) for any given activity. Thus, a user has access to the workbenches belonging to the activities which he is authorized to carry out.

Moreover, the system controls, by means of a built-in workflow module 5, the order in which activities can be carried out. At any given moment, the user has access for instance through a menu to all the activities which he can carry out at that particular point of time. The user can check the status of the various activities: which have been carried out, which can now be carried out, and which cannot (yet) be carried out?

Any activity is further subdivided into tasks, usually between 1-5 tasks per activity. Obviously, all the tasks belonging to one activity must be carried out by the same end-user role. When describing tasks the analyst notes which documents are used as input/output from and to the process. In some cases a task may produce only part of a document, e.g. a specific chapter for a manual.

Tasks usually involve language and documents and are often formulated (described) in specifically language-skill terms: *summarize the patient's wartime trauma, send a fax changing the appointment, report on the applicant's medical condition, read the applicant's supporting letter, etc.* Each task has potentially unique software support requirements.

The system contains unique language tasks support applications. These are step-by-step structure tools which help such language tasks as report writing, describing steps in a procedure, summarizing, inserting quotations, inserting and collecting references to other literature, and describing selection criteria for industrial materials, as shown in figure 1c. These tools are based on the latest linguistic insights into structured reading and writing strategies. They collect text in the form of step-by-step sub-sections which they gather into one document.

Tasks within one activity can be carried out in any order whereas the workflow module controls only at the level of activities.

Task-forms (screens) may be displayed to the user through the display

11.

A user can work at any one of a number of documents which are in production. They are stored in the dossier database 8 and a user may access them through e.g. a file option in a menu bar.

5 Preferably, a user can communicate with other users in the system by sending electronic messages over a network (for instance a local network or a wide-area network) or by sending a fax or generating a letter through communication module 13.

Now, with reference to figure 4 possible activities and tasks for the  
10 seven sub-processes listed above will be briefly explained. It is to be noted that this explanation relates to a possible embodiment only and is not intended to limit the scope of the present invention. Documents may be standard documents, letters, medical reports, etc.

a. Intake (steps 01.01 and 01.02 in figure 4)

15 A user can open an existing document through a file option in a menu bar. However, he can also change the identification of an old document in order to convert it into a new document, e.g. by making minor changes. He can also formulate a problem requiring the creation of a new document in order to obtain management authorization (an electronic signature) for its  
20 creation.

Browsing through the publication database 9, which includes all existing documents, enables the user to search for appropriate information, or to check its existence or non-existence.

The structure of a new document can be designed by selecting the  
25 appropriate content-structural items (including fixed configurations) from a semantic tree stored in the semantic database 7.

b. Collect information (steps 02.01 - 02.03 in figure 4)

30 Texts or objects (technical drawings (CAD), tables, graphs, formulae, or images) which are considered useful as input for the document being created can be imported into the dossier, which also serves to collect the ingoing and outgoing correspondence involved in the process.

The contents of the dossier can be viewed at any time by authorized users. Viewing is achieved by starting up an appropriate editor.

A user can search semantic database 7 for text fragments and objects  
35 which have formed part of existing documents by searching on keyword labels, text strings or semantic elements. A semantic element is a certain form of a piece of data. A piece of data may be described in various forms in order to be (re-)usable in a variety of different documents. In an in-

surance claim process, for instance, a data element may be *wartime trauma*. This data element occurs as several semantic elements, e.g. as a two-page interview transcript in a social report, as a three-line summary used for instructing the examining physician - and subsequently re-used for all 5 internal reports and summaries - and as a list of points used for the documents listing the grounds for acceptance or refusal of the claim. Semantic elements are an established concept in the field of linguistics. The present system is unique in the sense that it applies them to computer-supported document creation.

10 The analysis database 14 and the design database 17 are stored in predetermined memory locations. These databases allow the task description and list of semantic elements to appear on an end-user's screen on display 11 to be exported as a script file (used for generating programmable computer code), which is then imported into a screen editor. This allows the 15 designer to generate the list of semantic elements to appear on screen. The designer can then use, for instance, a computer mouse 2 to further arrange the items on screen and to check the validation of the semantic elements.

A user can also search the publishing database 9 for previously published material, for instance by using keyword searches, text searches or 20 Boolean searches.

The scanner 6 can be used to import paper documents as images or as digital text. In the latter case, use is made of optical character recognition (OCR).

CAD/CAM drawings can be imported where possible in CALS-compatible 25 formats.

Users can create annotated lists, summaries, and abstracts of relevant documents, and insert quotations by means of appropriate language task supports (see figure 1c), which are available per task, as explained above.

30 Facsimile letters and/or e-mail messages may be imported into the system through the communication I/O-module 13 and the internet module 18.

c. Edit (steps 03.01 -03.13 in figure 4)

A user can select a chapter he wants to work on, for instance, by clicking on an appropriate tab within the appropriate activity. Chapter 35 headings may be inserted automatically at the appropriate point in the text, but the wording can be edited by the user.

The user can insert sub-chapter and paragraph headings up to, e.g., four levels deep by clicking on appropriate options in the menu bar and the

tool bar.

The user may see the text page-by-page. Fonts and layout may be pre-defined. Navigating from page to page may, e.g., be done by clicking with the mouse 2. As an alternative, pages may be stacked or scrolled.

5 The tool bar offers strictly circumscribed possibilities with respect to bold type, underlining and italics, etc.

By means of printer 15 a user can print out previews of the document created at any time during sub-processes 1-5.

Preferably, the user can access the usual windows functionality of 10 cut, copy and paste, delete, search and replace, do and undo, at any moment.

He can embed and view objects in the text.

Objects can be positioned in the text by using pre-defined positions 15 and sizes. In the sub-process of editing, the following tasks may, for example, be supported by language task supports (figure 1c): describing required properties of materials, selecting materials and products, describing permitted materials and products, creating instructions, describing step-by-step procedures, etc.

By means of the terminology management modules 12 a user may be provided with the possibility of correcting spelling (module 12a) and certain 20 stylistic features, of checking and correcting terminology (modules 12b and 12c) and of providing dictionary references (modules 12d and 12e).

Both language task support modules 21 and language technology applications may be made available per task through the menu bar.

25 Domain-specific language and terminology may be collected and managed on-line by means of the terminology management modules 12.

d. Evaluate

Digital drafts may be sent by e-mail through the communication module 13 to appropriate specialists who can add comments digitally (for instance 30 by using electronic highlighting and sticky notes) and send digital versions back to the user (steps 03.14 - 03.17 in figure 4). After having received comments the user can view commented drafts stored in the dossier database 8 and consult them when editing the final draft (steps 03.01 - 03.13 again).

35 In the evaluation sub-process a language task support may enable the user to take minutes of discussions and telephone conversations. These minutes are, for instance, also stored in the dossier database 8.

In the evaluation sub-process, the user can use terminology manage-

ment modules 12 to check correctness and consistency.

In this sub-process the user can change the status of the document from manuscript to final draft, as required (step 03.18).

e. Translation (step 04 in figure 4)

5 Where required, the user can import a foreign-language version of the text to create a bilingual standard document.

Also a provision can be made for post-editing text in different languages.

f. Authorize (step 05.01 in figure 4)

10 A user who is provided with the role of manager can check the progress of each document in production by using the workflow module 5. Workflow modules are available on the market. The function of the workflow module 5 within the system of the present invention will be explained in detail later.

15 The manager can view the final draft and add comments to it, checking by means of an electronic check list. The manager may be allowed to send a document back to a former sub-process.

If he agrees to the final draft, the manager can place an electronic signature for approval. Then he can update the status of the document to 20 "approved", which may mean that it is ready for publication.

g. Publish (step 06 in figure 4)

The publishing database 9 contains a digital archive of approved documents. The user who is provided with the role of database administrator keeps track of and offers various versions (i.e. updates versions and 25 language versions) of documents.

Users having the role of authorized potential customers can browse through the publication database ... to view all standard documents and to search documents for specific information.

Documents or selections thereof can be published on demand, i.e., 30 when and as required, in various layouts, with various selections of information, on various media.

Preferably, when a new version of a published document is in preparation, the last published version is marked in a catalogue with a symbol indicating this. A published document which is older than five years, may 35 be marked in the catalogue as being administratively obsolete.

The blocks in the block diagram of figure 1a will be explained in detail but, first of all, the use of the "detector" as shown in figure 3 will be explained. The "detector" is associated with a method, for

instance, realised as a Lotus 123 (spreadsheet) application within the processor 1, which relates documents to business processes in order to evaluate their importance to the client's core business.

The detector as shown relates to a method which goes through the 5 following steps:

- a) the main processes of the business are entered on a diagram representing the value chain of activities carried out within a particular type of business. Until now, examples have been worked out for the process industry, banking and insurance;
- 10 b) documents are identified within the business, and are related to business processes in order to evaluate their importance to the client's core business;
- c) priorities and weightings are given to the documents and the processes;
- 15 d) these are calculated to present a prioritized list of the most important documents within the most important processes - these are candidates for a solution in accordance with the present invention.

The method may be realized as a spreadsheet application which can be used as follows:

- 20 a) during discussions with clients at top management level, processes and documents are entered into a spreadsheet;
- b) priorities are entered and calculated;
- c) charts and lists are produced for presentation back to decision-makers within the client company.

25 The object may be to create documents like letters, transcripts of telephone conversations, memo's, medical and scientific reports, protocols, brochures, articles, etc.

On the other hand, a more reduced number of very substantial documents, e.g. technical manuals, engineering and quality standard documents, 30 installation and maintenance guides, quality guidelines, etc. may be created.

Once a process has been identified by the "detector" the method may be continued by a "quick scan" method comprising the following steps:

- 35 a) the process detected is broken down into its component activities, which are listed;
- b) each activity is then assessed according to a number of parameters, including roles, cost, process and cycle time, customer value added, technology, linguistic complexity and document input/output;

c) these values are calculated to compare cost and cycle time, cost and customer value-added, complexity and process time, etc.

These steps may be realized as a spreadsheet application which can be used as follows:

5 a) during workshops with clients at management level, activities are listed and discussed;

b) comparisons and calculations are performed;

c) charts and overviews are produced for presentation back to decision-makers within the client company.

10 The analysis database 14 is a complex database containing the following:

a) a brief general description of the document production process, including the technical hardware and software environment in which it is carried out;

15 b) a list of sub-processes and a brief description of each;

c) a list of the activities, ordered per sub-process and a brief description of each, including the roles of the persons carrying this out, the triggers necessary to start the activity (e.g. availability of certain data) and the hardware and software currently available to support it;

20 d) a breakdown of the activities into tasks, including a brief description of each;

e) a list of all documents involved in the process (input, referenced or output), of whatever type, related to tasks wherever possible, but to a higher level (activities, sub-process) where this is not possible;

25 f) a list of the most important semantic elements used to build up the documents, related to each document (and thus, indirectly, to the tasks); with a description of each, including relational links to other semantic elements and to available data models;

30 g) a list of roles of users, including a brief profile of each role.

The analysis database application functions as follows:

a) workshops are held with a number of different groups at management level to elicit information at the most general level - much of the input during the quick scan method will be re-used and re-evaluated here;

b) participant interviews are conducted with a wide variety of end-users to elicit more detailed information and to collect documents;

c) documents are analyzed in depth and described by computer linguists;

- d) semantic elements are extracted from the documents and described;
- e) the various analyses are validated by being presented back to the clients.

The design database 17 is broadly similar in structure to the analysis database, but it contains much more explicit information at task and document level.

The design database 17 starts off as a copy of the analysis database 14, but it is then altered as follows:

- a) the process is compared to a generic process model for the particular document production process;
- b) this generic process is then customized for the client;
- c) the sub-processes and activities are streamlined to exclude redundant or unnecessarily lengthy or repetitive activities, and especially those activities which can be completely automated or at least extensively supported by automated tools;
- d) the number of roles is reduced to a minimum, and roles are matched to activities in such a way that an activity is specifically related to one particular role;
- e) the number of documents is reduced to a minimum;
- f) the semantic elements required to produce the documents are reduced to a minimum, and particular attention is paid to the re-usability of particular pieces of text, diagrams, formulae, etc.
- g) the task-forms (screens) required to handle the documents (input, reference and output) are described in terms of the necessary edit fields, buttons and other required controls;
- h) the controls on the task-forms are linked to the semantic elements in the database and their behaviour is described (defaults, active/non-active, press, edit, read/write, etc);
- i) each task description lists the necessary software applications, especially the language technology application stored in language technology modules 12, 20, 21 the communications technology and the external software applications which may be accessed through I/O module 13 and which are needed to support the particular task.

The design database 17 now contains all the information input necessary for the building of the client's customized application.

The system has now broken down the language-intensive activities into language tasks and is able to support each of these with a variety of software applications whereby the end-user is enabled to produce texts in such

a way that the content is highly structured, the resulting text is highly re-usable, the terminology and extra-textual references consistent and the style is adapted to the potential target group of the text.

Moreover, the system for supporting language-intensive business processes according to the invention is now customizable in such a way that the process structure implemented for a particular client reflects both the generic process-type and the specific process organization preferred by the client.

The design database 17 is now able to export the information required for each task-form as a script (i.e. a form of language which can be automatically converted into a computer language, e.g. C++). The processor 1 is provided with a task-form editor. The script is picked up by the task-form editor, which converts it into an object class (in the terminology of C++ object-oriented software engineering). This then automatically creates the task-form and puts it into the task-form library 10.

The task-form library 10 is the total stock of screens (task-forms) which are available to the end-users. An activity is divided into, for instance, 5 tasks. Each task is supported by a different task-form (screen), which has the appearance of a tab. All the task-forms are stored in one place, and are invoked when necessary by the processor 1. It is possible that a task-form is employed by more than one task. Task-forms serve to present information to the user and to enter or select information by the user. Task-forms can check the integrity, consistency or validity of particular information (e.g. format of dates/codes/numbers, relation between dates, etc.).

The logon/logoff module 3 regulates access to various activities. When a user logs on this module establishes the level of access allowed by the user, i.e., which tasks he is allowed to carry out throughout the system. The logon/logoff module 3 uses the data stored in the user profile data base 16 to recognize the user and the role(s) belonging to the user. Each user may have more than one role. The user can logon/logoff per individual role.

The processor 1 integrates the total architecture, and regulates the functionality available per user and per activity/task. The processor 1 governs among others:

- the different workbenches available for each activity;
- which screens (task-forms/tabs) are displayed to the user;
- which language technology applications are available per task;

- which other applications (scanner, financial modules, etc.) are available;
- the information from the workflow module 5;
- the storage of information in the various databases;
- 5 - the link, via the communication module 13, to other systems, e.g., other computers at the client's site;
- the access to the dossier database 8 (folder application).

The interchanger 4 provides an interface between the processor 1 and any workflow module 5. It can link the processor 1 to a particular commercially available piece of software. If a user wishes to use workflow software which is generally available on the market, this can easily be integrated into the system by means of the interchanger 4.

In cases when such a commercially available package is used, it generally replaces the system workflow module 5, tracking database 19 and the logon/logoff module 3 but not, of course, the interchanger 4 itself which makes the integration possible.

The workflow application software:

- governs authorization of users (identified in logon/logoff module 3) to carry out particular activities;
- 20 - lists possible activities and sequence in which they may be carried out;
- lists "triggers" which enable activities to be done (e.g. a certain piece of information has become available, so a certain letter can be sent);
- keeps tracking information in the tracking database 19, e.g. how much process time has been spent, which information has been stored in the databases, etc. There are many of these workflow applications available on the market.

The scanner 6 is an example of an external application which has been integrated into the system according to the invention. The processor 1 governs when an end-user can/must scan during the process (e.g. when receiving incoming paper correspondence). He uses the scanner and the appropriate third party software to scan the letter in and store it in the dossier database 8.

In order to link the system's task-forms to the scanner software the processor 1 is provided with specially developed scanner application software. When a task entails scanning, this application software starts up the scanner software and hardware of the scanner 6.

The processor 1 is also provided with a piece of software which may be called a folder application and enables the user to manage all kinds of

information (e.g. text/word processor files, 'objects' like pictures, drawings, graphs, images, but also movies and sound files, etc.) related to the case. Through the folder application the user gets access to the dossier database 8 belonging to a case.

5 The processor 1 may include an object server for storing and retrieving objects in a networked environment. These objects can be displayed on the display 11.

10 The entity in which the end-user actually sees objects is called a dossier. The dossier can be opened from the menu bar. It contains a series of tabs, classified according to the type of process - typically: correspondence, financial information, background info., etc. The actual tabs are completely customizable. All documents belong to and are stored with the case (document being produced).

15 Under a dossier there are lists of the documents and objects. When the user clicks on the name of a document, the appropriate application (e.g. AutoCad, WordPerfect, Corel Draw) is started up so that the object can be viewed and modified. The dossier is stored in the publishing database 9 once the document it belongs to has been published.

20 The semantic database 7 is a relational database which stores all semantic elements. Technically speaking, semantic elements (SE's) are like higher level entities which encapsulate more traditional features of data elements. A specific data element (a meaningful element e.g. a paragraph about a certain topic) appears in different forms depending on the context of use. Each version of a piece of data is a semantic element: a unique 25 combination of form and meaning. The semantic database 7 is gradually filled during the intake activities of the process, and the semantic elements are (re)used when producing documents.

30 The publishing database 9 contains all documents which need to be made available to external or internal users. It contains the documents in the following form:

- a) a list of semantic elements required to generate the document;
- b) the objects the document must contain (CAD drawings, tables, audio, etc);
- c) formatting information for printing (e.g. a desktop publishing application style sheet);
- d) an SGML format of each document for compatibility with international standards;
- e) other formats for electronic publishing, Internet publishing, etc.

Language task support applications are small applications, which may be continuously developed and all have as their general functionality:

- a) they are all started from the menu bar uniquely configured to accompany a particular task-form;
- 5 b) the language task to be performed is split into a number of steps; described through a language skills methodology (derived from communicative language teaching theory in the field of applied linguistics);
- 10 c) each of the language task support modules 21 consists of a number of screens, each with specific input fields, radio buttons and other screen components, ordered in such a way that the task is automatically carried out correctly;
- 15 d) the completed text is inserted as a software object into the document being created, and is stored as one or more semantic elements - this means that it cannot be manipulated as normal text; when selected, the text originating from a language task support module can only be modified by starting up the original language task support module;
- 20 e) the resulting structured text can be linked to style sheets to produce an output document with a clear, appropriate layout and visual structure, thus aiding comprehensibility.

Examples of language task support applications include summary writing, writing a medical report, inserting quotations into a text, describing materials, enumerating a procedure, etc.

Thus, linguistic structures are used to support document production processes. A linguistic structure of a text is defined as the way in which its content is logically ordered in a clearly understandable way, e.g. as a chronological account, as a logical ordering of successive steps, as a typology or hierarchy, as a sequence of questions and answers, etc.

A clear linguistic structure employs natural language to point up this logical ordering by using enumeration *Firstly*, ..., by indicating time sequence *After that*, *you* ..., by indicating logical relationships *Contrary to what might be expected*, ..., etc.

The use of correct paragraphing, punctuation, well-designed layout and functional illustrations can enhance the clarity of the linguistic structure. When used for this purpose, these aspects are known as *paralinguistic features*.

Re-usable texts can be inserted by the end-user in a variety of different documents, aimed at a variety of different target groups. Re-usabil-

ity depends on the style being appropriate, on the total length, the degree of abstraction, the linguistic structure, the type of terminology used, the ability to retrieve them in an electronic database, etc.

Re-usability saves time and leads to more consistent documents with 5 predictable linguistic structure, leading to greater ease of reading (readability).

The system also comprises a checklist application and checklist management. Edit fields on the task-forms frequently have a small application attached called the checklist application. When a user clicks on a screen 10 button, a small window opens up with a list of key-words that the user can choose from. The topics covered by these keywords relate to the semantic element which can be entered in that particular edit field. Keywords do not have to occur in the text of the semantic element, but they do cover topics dealt with.

15 Selected keywords are stored with the semantic element in the semantic database 7. The user can use them to search for certain information, e.g. all semantic elements in the semantic database 7 marked with the keyword *viscosity*.

Keywords can also be used by a user having the role of "terminology 20 manager" to collect a body of texts (a corpus) for analysis and subsequent terminology extraction. To this end, checklist management is provided. Its purpose is to create or maintain consistency in keywords and use of keywords within the checklist application. The checklist management activity is only available to the "terminology manager". The checklist management 25 activity can be thought of as consisting of the following steps: collection of checklists, review of how existing keywords are used (when and where), review of keywords added, monitoring possible conflicts, team-discussions, updating the standard keywords in checklists, and elimination of inconsistencies.

30 The system provides the possibility of integrating language technology applications of third parties. Thus, extensive integration of language technology in the document production process is possible. The language technology applications can be opened from the menu bar, and are available for those task-forms where they may be required.

35 When these applications have been acquired from third parties they may be offered on a "pay-as-you-use" on demand basis. A timing system counts how long the user uses the application, and bills him periodically. The owners of the software receive a royalty from the system for this

usage. Examples are: spelling checkers, style checkers, electronic thesauruses, automatic translation software, brainstorm software (for generating and describing ideas), and electronic dictionaries. Communications with remotely located applications may be done through the communication module

5 13.

The system may also be provided with a number of language technology applications not available elsewhere on the market. Particularly noteworthy is the provision of a "terminology management cycle". This performs the following functions:

- 10 a) a corpus of texts is built up by the "terminology manager", using the checklist management application;
- b) this corpus is analyzed, using the concordance module 12f (this is a piece of software available on the market);
- c) terms are identified and extracted;
- 15 d) document authors use a dynamic term collection application to indicate new terms in the texts they have just written, collect them and compare them to the terms already in the database;
- e) terms from (c) and (d) are submitted to an expert panel for approval;
- f) terms are added to the database;
- 20 g) terms are exported to a glossary application (monolingual explanation of terms and lists of equivalence across different domains);
- h) terms are exported to enhance the spelling checker 12a (this is a piece of software available on the market);
- i) terms are automatically generated as monolingual or multilingual look-up dictionaries, unique to the client's organization;
- 25 j) terms are exported to a term checker which checks texts for correct (consistent) use of terms, suggesting correct terms where incorrect ones have been used.
- k) the thus enhanced language technology applications improve the quality of the texts produced, thus improving the quality of the corpus to be analyzed - the more the terminology management cycle is used, the better it functions, and the better the quality of the texts.

If required, third party software applications (e.g. spreadsheets, operating systems, workflow software) may be integrated into specific 35 customized versions of the system. Software is provided which ensures the necessary integration on a technical level. The end-user is unaware of this.

Moreover, specially developed software is provided which enables the

user to search in the semantic database 7 and/or the publishing database 9 for documents and/or semantic elements on the basis of keywords in the checklist application and document attributes (title, number, date, author, etc.). The underlying functionality is customized for each particular client application.

Thus, while using the system, the end-user builds up lists of terminology, text-components, paralinguistic objects (audio-visuals, images, drawings). These serve to enrich the analysis and enhance the applications available. Thus, the more the system is used, the more powerful it gets.

The total integration of the diverse systems, the unique support of language tasks and the cyclical application of corpus analysis and terminology management create a synergy which surpasses the possibilities of the sum of the various pieces of third party software.

It will be evident to persons skilled in the art that the functional blocks of the system as presented in figure 1a need not be implemented on separate hardware units. Some may be integrated on one single chip, as required.

## Claims

1. Document producing support system comprising:

- input means (2) for receiving manual input from a user;
- 5 - display means (11);
- a user profile database (16) storing user profiles;
- a logon/logoff module (3) connected to said user profile database (16) for identifying roles of users based on said user profiles;
- a semantic database (7) storing semantic elements, i.e., pieces of data with pre-defined contents and meanings;
- a dossier database (8) for storing drafts of documents;
- a publishing database (9) for storing final versions of documents;
- a work flow module (5) arranged for at least controlling which roles of users are entitled to carry out specific activities and which activities may be carried out at a specific moment, as well as tracking which activities have already been carried out and how much process time has already been spent within any of the activities, activities being defined as sets of tasks allowed to be carried out by any individual user depending on his role;
- 20 - language technology modules (12, 20, 21) for terminology control, machine-aided translation and language task support;
- a processor (1) coupled to and arranged for controlling said input means (2), said logon/logoff module (3), said work flow module (5), said display means (11), said language technology modules (12, 20, 21), and said semantic database (7), said dossier database (8) and said publishing database (9).

2. System according to claim 1 furthermore comprising a task form library database (10) storing data relating to a set of task forms, each task form being associated with at least one task, said processor (1) being coupled to said third memory means for displaying said task forms on said display means (11).

3. System according to any of the preceding claims furthermore comprising an analysis database (14) storing the following data:

- a) a brief general description of the document production process, including technical hardware and software environment in which it is carried out;

- b) a list of sub-processes and a description of each;
- c) a list of activities, ordered per sub-process and a description of each, including roles of persons carrying this out, triggers necessary to start any of the activities and the hardware and software currently available to support it;
- 5 d) a breakdown of the activities into tasks, including a description of each;
- e) a list of all documents involved in the process, related to the tasks;
- 10 f) a list of semantic elements used to build up the documents, related to each document, with a description of each, including relational links to other semantic elements and to available data models;
- g) a list of roles of users, including a brief profile of each role.

15 4. System according to any of the preceding claims furthermore comprising an interchanger (4) coupled between said workflow module (5) and said processor (1) for providing an interface between a commercially available workflow module and the processor (1).

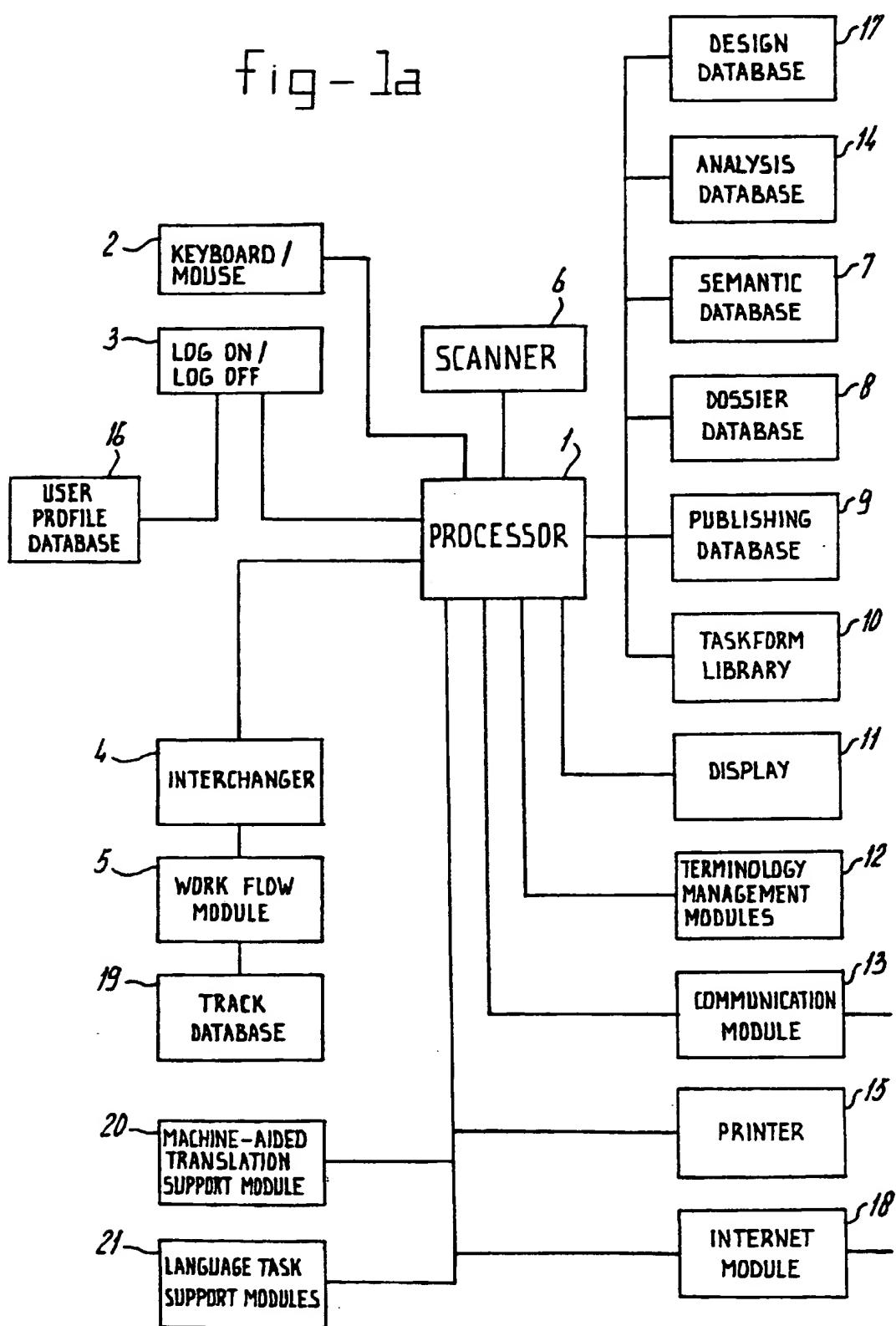
20 5. System according to any of the preceding claims furthermore comprising a scanner (6) connected to said processor (1) for scanning paper documents and transferring the content to said processor (1).

25 6. System according to any of the preceding claims furthermore comprising an I/O module (13) for coupling the system to external systems.

7. System according to any of the preceding claims furthermore comprising an internet module (18) for providing documents to internet.

1/5

fig - 1a



2/5

Fig - 1b

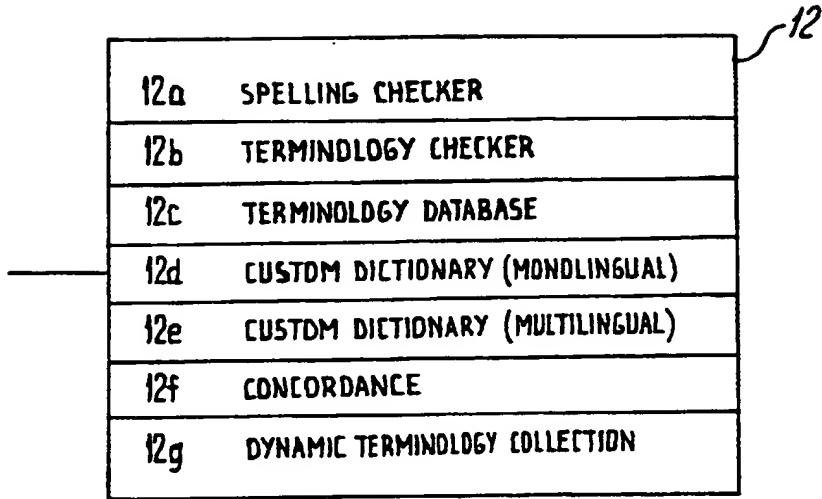
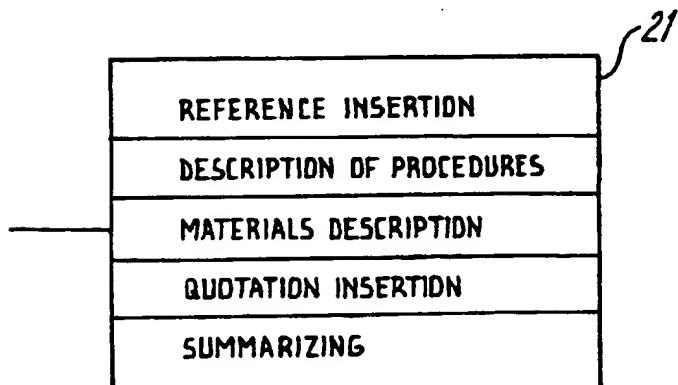
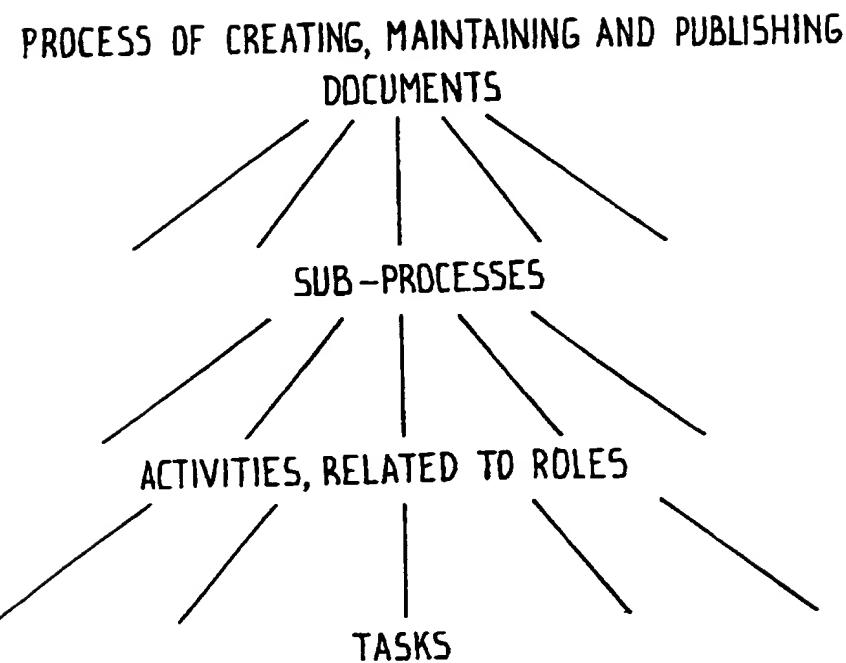


Fig - 1c



3/5

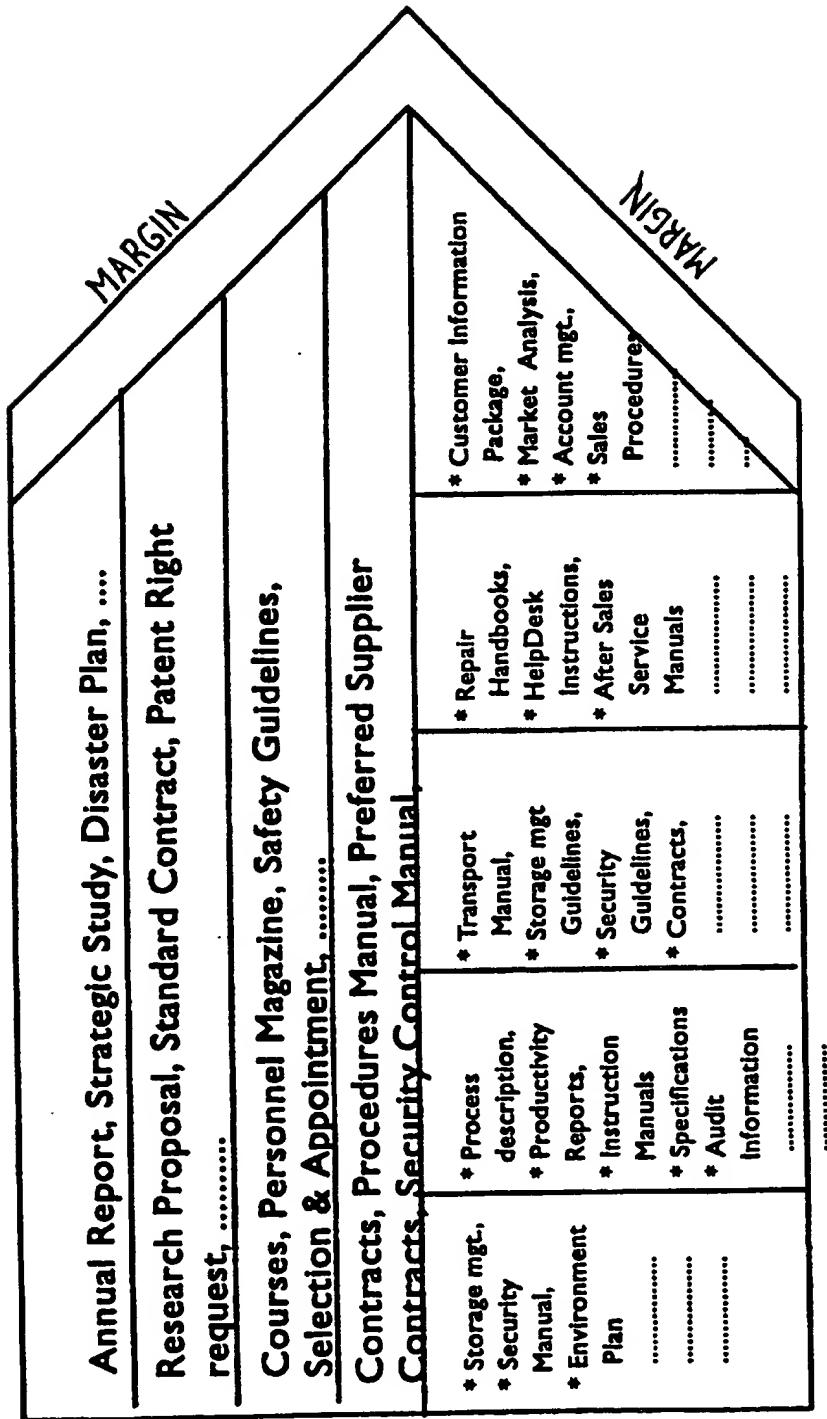
fig-2



4/5

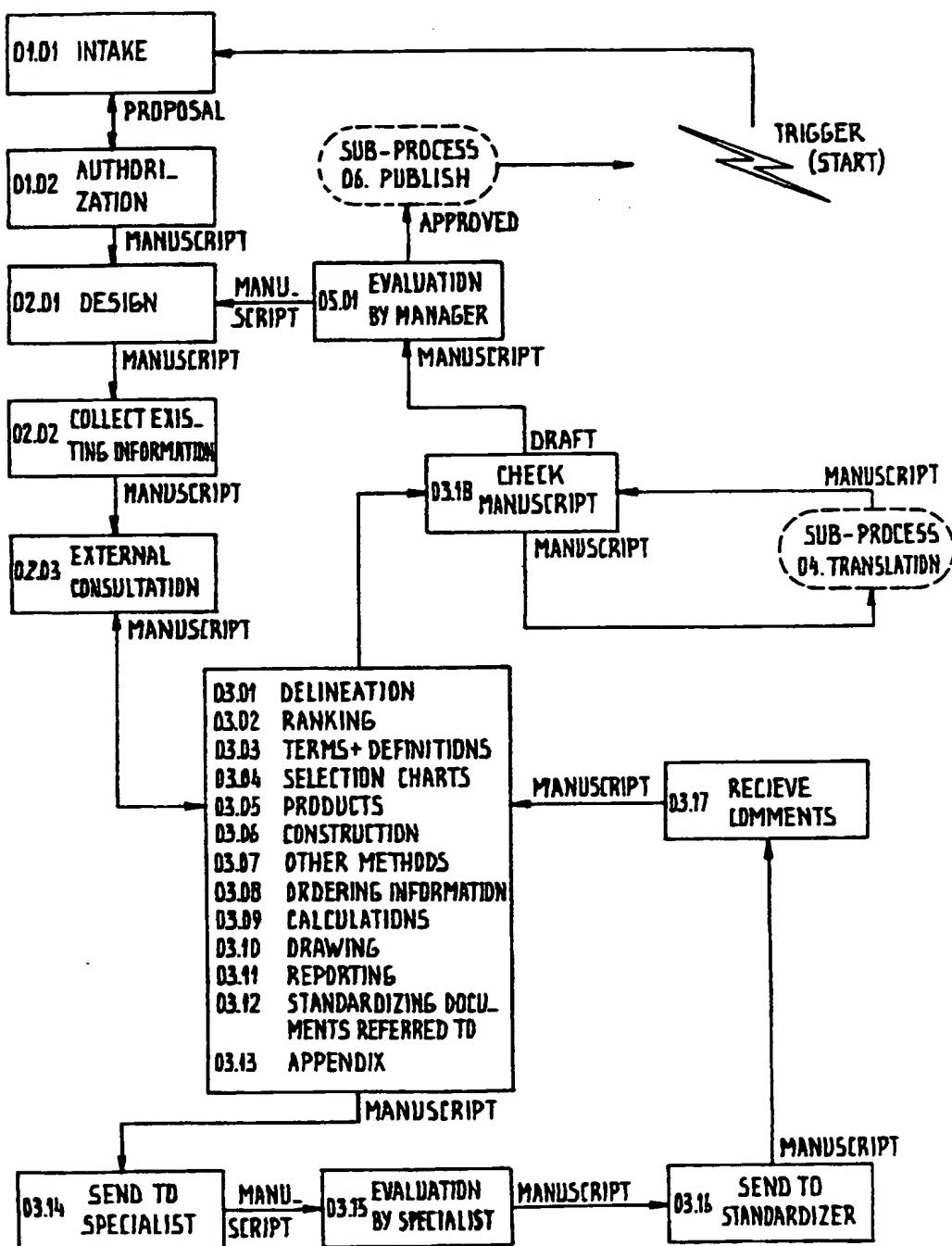
# Detector

Fig - 2



5/5

fig - 4



## INTERNATIONAL SEARCH REPORT

International Application No  
PCT/NL 96/00273

A. CLASSIFICATION OF SUBJECT MATTER  
IPC 6 G06F17/28 G06F17/60

According to International Patent Classification (IPC) or to both national classification and IPC

## B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)  
IPC 6 G06F

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

## C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	WO,A,94 18620 (ACTION TECHNOLOGIES INC) 18 August 1994 see abstract; claims 1-3; figure 2 ---	1-7
Y	EP,A,0 697 666 (POLYDOC N V) 21 February 1996 see column 4, line 15 - column 5, line 15 ---	1-7
A	EP,A,0 603 513 (MOTOROLA INC) 29 June 1994 see abstract; claims 1-10; figure 1 ---	1,5
A	EP,A,0 472 786 (ITT) 4 March 1992 see page 9, line 5 - line 55 see page 44, line 32 - page 45, line 55 ---	1-4
A	WO,A,94 06086 (CATERPILLAR INC) 17 March 1994 see abstract; claims 7,9 ---	1
	-/-	

Further documents are listed in the continuation of box C.

Patent family members are listed in annex.

## \* Special categories of cited documents :

- \*A\* document defining the general state of the art which is not considered to be of particular relevance
- \*E\* earlier document but published on or after the international filing date
- \*L\* document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
- \*O\* document referring to an oral disclosure, use, exhibition or other means
- \*P\* document published prior to the international filing date but later than the priority date claimed

- \*T\* later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
- \*X\* document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
- \*Y\* document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.
- \*A\* document member of the same patent family

1

Date of the actual completion of the international search

7 February 1997

Date of mailing of the international search report

14.02.97

Name and mailing address of the ISA

European Patent Office, P.B. 5818 Patentlaan 2  
NL - 2230 HV Rijswijk  
Tel. (+31-70) 340-2040, Tx. 31 651 epo nl,  
Fax (+31-70) 340-3016

Authorized officer

Suendermann, R

## INTERNATIONAL SEARCH REPORT

Inte  
al Application No  
PCT/NL 96/00273

## C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	EP,A,θ 692 765 (CANON KK) 17 January 1996 see abstract; claims 1-31 ----	1
A	IBM TECHNICAL DISCLOSURE BULLETIN, vol. 34, no. 7B, December 1991, NEW YORK, US. pages 499-500, XP002024781 ANONYMOUS: "Information Development Data Base." see abstract -----	1

# INTERNATIONAL SEARCH REPORT

Information on patent family members

Int'	inal Application No
	PCT/NL 96/00273

Patent document cited in search report	Publication date	Patent family member(s)	Publication date	
WO-A-9418620	18-08-94	AU-A- EP-A-	6133594 0686282	29-08-94 13-12-95
EP-A-0697666	21-02-96	NL-A-	9401314	01-03-96
EP-A-0603513	29-06-94	JP-A-	6231139	19-08-94
EP-A-0472786	04-03-92	AT-T- DE-D-	135477 69025935	15-03-96 18-04-96
WO-A-9406086	17-03-94	AU-A- CA-A- CN-A- EP-A- FI-A- JP-T-	5087993 2141221 1083952 0658259 951010 8501166	29-03-94 17-03-94 16-03-94 21-06-95 03-03-95 06-02-96
EP-A-0692765	17-01-96	JP-A- JP-A-	8006952 8006935	12-01-96 12-01-96